

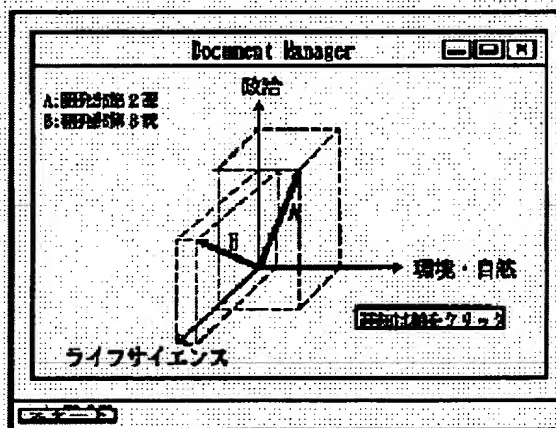
DOCUMENT PROCESSOR, STORAGE MEDIUM STORING DOCUMENT PROCESSING PROGRAM AND DOCUMENT PROCESSING METHOD

Patent number: JP11045286
Publication date: 1999-02-16
Inventor: NOMURA NAOYUKI; MITOBE KATSUHIKO
Applicant: JUST SYST CORP
Classification:
- **International:** G06F17/30; G06F17/27
- **European:**
Application number: JP19970217154 19970728
Priority number(s): JP19970217154 19970728

Report a data error here

Abstract of JP11045286

PROBLEM TO BE SOLVED: To provide a document processor, a storage medium storing a document processing program and a document processing method capable of visualizing and expressing the preference of a user and recognizing difference and secular change by the user. **SOLUTION:** A GP(group personalizing) matrix for which one of the user and a processing important word obtained from an appearing frequency or the like in the processing document in the past of the user is turned to a row, the other is turned to a column and the importance of the respective processing important words for the user is turned to an element value is obtained. Then, a preference document vector is obtained by shifting a document vector for which the importance of the important word of a reference document is an element by the GP matrix, the respective elements (importance) of the preference document vector are totaled for respective fields, the importance $F(X)$ for the respective fields is calculated and the preference document vector is expressed on three dimensions for which the three fields A, B and C of the high importance $F(X)$ for the respective fields are respective axes.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-45286

(43) 公開日 平成11年(1999) 2月16日

(51) Int.Cl.⁶

識別記号

F I

G 0 6 F 17/30
17/27

G 0 6 F 15/403 3 4 0 A
15/20 5 5 0 F
15/40 3 7 0 A
15/403 3 5 0 C

審査請求 未請求 請求項の数12 F D (全 12 頁)

(21) 出願番号 特願平9-217154
(22) 出願日 平成9年(1997) 7月28日

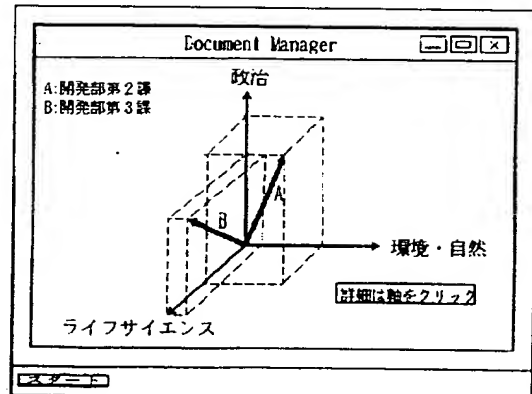
(71) 出願人 390024350
株式会社ジャストシステム
徳島県徳島市沖浜東3-46
(72) 発明者 野村 直之
徳島県徳島市沖浜東3丁目46番地 株式会
社ジャストシステム内
(72) 発明者 水戸部 勝彦
徳島県徳島市沖浜東3丁目46番地 株式会
社ジャストシステム内
(74) 代理人 弁理士 川井 隆 (外1名)

(54) 【発明の名称】 文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法

(57) 【要約】

【課題】 ユーザーの嗜好を視覚化して表現し、ユーザーによる差異や経時的変化を認識できる文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法を提供すること。

【解決手段】 ユーザーと、前記ユーザーの過去の処理文書中における出現頻度等から取得した処理重要語の一方を行、他方を列とし、前記ユーザーに対する前記各処理重要語の重要度を要素値とするG P行列を取得する。そして基準文書の重要語の重要度を要素とする文書ベクトルをG P行列によりシフトさせて嗜好文書ベクトルを取得し、嗜好文書ベクトルの各要素(重要度)を分野別に総計し、分野別重要度F(X)を算出し、分野別重要度F(X)の高い3分野A、B、Cを各軸とする3次元上に、嗜好文書ベクトルを表現する。



1

【特許請求の範囲】

【請求項1】 過去に処理された文書から、ユーザーと、前記ユーザーの嗜好を表す複数のキーワードの一方を行、他方を列とし、前記ユーザーに対する前記各キーワードの重要度を要素値とするGP行列を取得するGP行列取得手段と、

前記GP行列を視覚化するGP行列視覚化手段と、を具備することを特徴とする文書処理装置。

【請求項2】 前記GP行列視覚化手段は、前記GP行列から前記キーワードの重要度を要素値とするGPベクトルを取得し、このGPベクトルを n ($n \geq 2$)次元化して表示することを特徴とする請求項1に記載の文書処理装置。

【請求項3】 文書の特徴付ける文書ベクトルを取得する文書ベクトル取得手段を備え、

前記GP行列視覚化手段は、前記文書ベクトルを前記GP行列を用いてシフトさせて前記GPベクトルを取得し、前記文書ベクトルと前記GPベクトルとを表示することを特徴とする請求項2に記載の文書処理装置。

【請求項4】 文書の特徴付ける文書ベクトルを取得する文書ベクトル取得手段を備え、

前記GP行列視覚化手段は、同一のユーザーに対する前記GPベクトルの経時的变化を表示することを特徴とする請求項2または請求項3に記載の文書処理装置。

【請求項5】 前記GP行列視覚化手段は、複数の前記ユーザーそれぞれについての前記GPベクトルを同時に表示するものであることを特徴とする請求項2から請求項4のうちのいずれか1の請求項に記載の文書処理装置。

【請求項6】 過去に処理された文書から、ユーザーと、前記ユーザーの嗜好を表す複数のキーワードの一方を行、他方を列とし、前記ユーザーに対する前記各キーワードの重要度を要素値とするGP行列を取得するGP行列取得機能と、

前記GP行列を視覚化するGP行列視覚化機能と、をコンピュータに実現させるためのコンピュータ読みとり可能な文書処理プログラムが記憶された記憶媒体。

【請求項7】 前記GP行列視覚化機能は、前記GP行列から前記キーワードの重要度を要素値とするGPベクトルを取得し、このGPベクトルを n ($n \geq 2$)次元化して表示することを特徴とする請求項6に記載の文書処理プログラムが記憶された記憶媒体。

【請求項8】 文書の特徴付ける文書ベクトルを取得する文書ベクトル取得機能を備え、

前記GP行列視覚化機能は、前記文書ベクトルを前記GP行列を用いてシフトさせて前記GPベクトルを取得し、前記文書ベクトルと前記GPベクトルとを表示することを特徴とする請求項7に記載の文書処理プログラムが記憶された記憶媒体。

【請求項9】 文書の特徴付ける文書ベクトルを取得す

2

る文書ベクトル取得機能を備え、

前記GP行列視覚化機能は、同一のユーザーに対する前記GPベクトルの経時的变化を表示することを特徴とする請求項7または請求項8に記載の文書処理プログラムが記憶された記憶媒体。

【請求項10】 前記GP行列視覚化機能は、複数の前記ユーザーそれぞれについての前記GPベクトルを同時に表示するものであることを特徴とする請求項7から請求項9のうちのいずれか1の請求項に記載の文書処理プログラムが記憶された記憶媒体。

【請求項11】 過去に処理された文書から、ユーザーと、前記ユーザーの嗜好を表す複数のキーワードの一方を行、他方を列とし、前記ユーザーに対する前記各キーワードの重要度を要素値とするGP行列を取得し、前記GP行列を視覚化することを特徴とする文書処理方法。

【請求項12】 前記GP行列から前記キーワードの重要度を要素値とするGPベクトルを取得し、このGPベクトルを n ($n \geq 2$)次元化して表示することにより前記GP行列を視覚化することを特徴とする請求項11に記載の文書処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法に関し、更に詳細には、ユーザーの嗜好を視覚化して表現し、ユーザーによる差異や経時的变化を認識できる文書処理装置、文書処理プログラムを記憶した記憶媒体及び文書処理方法に関する。

【0002】

【従来の技術】従来の文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法による文書処理においては、文書をベクトル化して文書ベクトルとして表すことが行われている。この文書ベクトルは、それぞれの文書におけるキーワードの出現回数等を要素として取得され、各文書の特徴付けるものとなっているので、文書の検索・分類等を行う場合の目安として有用である。

【0003】

【発明が解決しようとする課題】しかし、同一の文書でも、例えば営業用や技術資料用等の利用目的その他のユーザーの嗜好が異なると、重要部位等に差異が生じる。また、同一のユーザーであっても、その嗜好は経時に変化する場合がある。そのため、従来より、文書の特徴を文書ベクトルとして表すのと同様に、ユーザーの嗜好を視覚化して表現し、ユーザーによる差異や経時的变化を認識できる技術が望まれていた。

【0004】本発明は、上述のような課題を解決するためになされたもので、ユーザーの嗜好を視覚化して表現し、ユーザーによる差異や経時的变化を認識できる文書

50

処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法を提供することを目的とする。

【0005】

【課題を解決するための手段】請求項1に記載の発明は、図15に示すように、過去に処理された文書から、ユーザーと、前記ユーザーの嗜好を表す複数のキーワードの一方を行、他方を列とし、前記ユーザーに対する前記各キーワードの重要度を要素値とするGP行列を取得するGP行列取得手段101と、前記GP行列を視覚化するGP行列視覚化手段102と、を具備する文書処理装置を提供することにより、前記目的を達成する。請求項2に記載の発明は、図15に示すように、請求項1に記載の文書処理装置において、前記GP行列視覚化手段102は、前記GP行列から前記キーワードの重要度を要素値とするGPベクトルを取得し、このGPベクトルを n ($n \geq 2$)次元化して表示する文書処理装置を提供することにより、前記目的を達成する。請求項3に記載の発明は、図16に示すように、請求項2に記載の文書処理装置において、文書の特徴付ける文書ベクトルを取得する文書ベクトル取得手段103を備え、前記GP行列視覚化手段102は、前記文書ベクトルを前記GP行列を用いてシフトさせて前記GPベクトルを取得し、前記文書ベクトルと前記GPベクトルとを表示する文書処理装置を提供することにより、前記目的を達成する。請求項4に記載の発明は、図16に示すように、請求項2または請求項3に記載の文書処理装置において、文書の特徴付ける文書ベクトルを取得する文書ベクトル取得手段103を備え、前記GP行列視覚化手段102は、同一のユーザーに対する前記GPベクトルの経時的变化を表示する文書処理装置を提供することにより、前記目的を達成する。請求項5に記載の発明は、図15または図16に示すように、請求項2から請求項4のうちのいずれか1の請求項に記載の文書処理装置において、前記GP行列視覚化手段102は、複数の前記ユーザーそれぞれについての前記GPベクトルを同時に表示するものである文書処理装置を提供することにより、前記目的を達成する。請求項6に記載の発明は、図17に示すように、過去に処理された文書から、ユーザーと、前記ユーザーの嗜好を表す複数のキーワードの一方を行、他方を列とし、前記ユーザーに対する前記各キーワードの重要度を要素値とするGP行列を取得するGP行列取得機能201と、前記GP行列を視覚化するGP行列視覚化機能202と、をコンピュータに実現させるためのコンピュータ読みとり可能な文書処理プログラムが記憶された記憶媒体を提供することにより、前記目的を達成する。請求項7に記載の発明は、図17に示すように、請求項6に記載の記憶媒体において、前記GP行列視覚化機能202は、前記GP行列から前記キーワードの重要度を要素値とするGPベクトルを取得し、このGPベクトルを n ($n \geq 2$)次元化して表示する文書処理プログラム

が記憶された記憶媒体を提供することにより、前記目的を達成する。請求項8に記載の発明は、図18に示すように、請求項7に記載の記憶媒体において、文書の特徴付ける文書ベクトルを取得する文書ベクトル取得機能203を備え、前記GP行列視覚化機能202は、前記文書ベクトルを前記GP行列を用いてシフトさせて前記GPベクトルを取得し、前記文書ベクトルと前記GPベクトルとを表示する文書処理プログラムが記憶された記憶媒体を提供することにより、前記目的を達成する。請求項9に記載の発明は、図18に示すように、請求項7または請求項8に記載の記憶媒体において、文書の特徴付ける文書ベクトルを取得する文書ベクトル取得機能203を備え、前記GP行列視覚化機能202は、同一のユーザーに対する前記GPベクトルの経時的变化を表示する文書処理プログラムが記憶された記憶媒体を提供することにより、前記目的を達成する。請求項10に記載の発明は、図17または図18に示すように、請求項7から請求項9のうちのいずれか1の請求項に記載の記憶媒体において、前記GP行列視覚化機能202は、複数の前記ユーザーそれぞれについての前記GPベクトルを同時に表示するものである文書処理プログラムが記憶された記憶媒体を提供することにより、前記目的を達成する。請求項11に記載の発明は、図19に示すように、過去に処理された文書から、ユーザーと、前記ユーザーの嗜好を表す複数のキーワードの一方を行、他方を列とし、前記ユーザーに対する前記各キーワードの重要度を要素値とするGP行列を取得301し、前記GP行列を視覚化する302ことを特徴とする文書処理方法を提供することにより、前記目的を達成する。請求項12に記載の発明は、図19に示すように、請求項11に記載の文書処理方法において、前記GP行列から前記キーワードの重要度を要素値とするGPベクトルを取得し、このGPベクトルを n ($n \geq 2$)次元化して表示することにより前記GP行列を視覚化する302文書処理方法を提供することにより前記目的を達成する。

【0006】

【発明の実施の形態】以下、本発明の文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法の好適な実施の形態について、図1から図10を参照して詳細に説明する。

(1)実施形態の概要

本実施形態では、ユーザーが過去の処理文書中における出現頻度等から、処理重要語およびこれらの処理重要度によりユーザーの嗜好を表すGP行列を取得する。そして基準文書の重要語の重要度を要素とする文書ベクトルをGP行列によりシフトさせて嗜好文書ベクトルを取得し、嗜好文書ベクトルの各要素(重要度)を分野別に総計し、分野別重要度 $F(X)$ を算出し、分野別重要度 $F(X)$ の高い3分野A、B、Cを各軸とする3次元上に、嗜好文書ベクトルを表現する。

【0007】(2)実施形態の詳細

図1は、本発明の文書処理装置の一実施形態であり、本発明の文書処理プログラムを記憶した記憶媒体の一実施形態の該プログラムが読み取られたコンピュータの構成を表したブロック図である。この図1に示すように、文書処理装置(コンピュータ)は、装置全体を制御するための制御部11を備えている。この制御部11には、データバス等のバスライン21を介して、入力装置としてのキーボード12やマウス13、表示装置14、印刷装置15、記憶装置16、記憶媒体駆動装置17、通信制

御装置18、および、入出力I/F19、および、文字認識装置20が接続されている。制御部11は、CPU111、ROM112、RAM113を備えている。ROM112は、CPU111が各種制御や演算を行うための各種プログラムやデータが予め格納されたリードオンリーメモリである。

【0008】RAM113は、CPU111にワーキングメモリとして使用されるランダムアクセスメモリである。このRAM113には、本実施形態による文書ベクトル取得処理を行うためのエリアとして、文書ベクトル取得の対象となる文書を格納する対象文書格納エリア1131、キーワード格納エリア1132、文書ベクトル格納エリア1134が確保され、また、GP行列取得処理を行うためのエリアとして、行列格納エリア1135、GP行列視覚化処理を行うためのエリアとして、GPベクトル格納エリア1138その他の各種エリアが確保されるようになっている。

【0009】キーボード12は、かな文字を入力するためのかなキーやテンキー、各種機能を実行するための機能キー、カーソルキー、等の各種キーが配置されている。マウス13は、ポインティングデバイスであり、表示装置14に表示されたキーやアイコン等を左クリックすることで対応する機能の指定を行う入力装置である。表示装置14は、例えばCRTや液晶ディスプレイ等が使用される。この表示装置14には、文書ベクトルを得る対象文書の内容や、本実施形態により取得されたGP行列が視覚化された嗜好文書ベクトル、等が表示されるようになっている。印刷装置15は、表示装置14に表示された文書や、記憶装置16の文書データベース164に格納された文書等の印刷を行うためのものである。この印刷装置としては、レーザプリンタ、ドットプリンタ、インクジェットプリンタ、ページプリンタ、感熱式プリンタ、熱転写式プリンタ、等の各種印刷装置が使用される。

【0010】記憶装置16は、読み書き可能な記憶媒体と、その記憶媒体に対してプログラムやデータ等の各種情報を読み書きするための駆動装置で構成されている。この記憶装置16に使用される記憶媒体としては、主としてハードディスクが使用されるが、後述の記憶媒体駆動装置17で使用される各種記憶媒体のうちの読み書き

可能な記憶媒体を使用するようにしてもよい。記憶装置16は、仮名漢字変換辞書161、プログラム格納部162、文書データベース164、文書ベクトルデータベース166、行列データベース168、図示しないその他の格納部(例えば、この記憶装置16内に格納されているプログラムやデータ等をバックアップするための格納部)等を有している。プログラム格納部162には、本実施形態における文書ベクトル取得処理プログラム、GP行列取得処理プログラム、GP行列視覚化処理プログラム等の各種プログラムの他、仮名漢字変換辞書161を使用して入力された仮名文字列を漢字混り文に変換する仮名漢字変換プログラム等の各種プログラムが格納されている。

【0011】文書データベース164には、仮名漢字変換プログラムにより作成された文書や、他の装置で作成されて記憶媒体駆動装置17や通信制御装置18から読み込まれた文書が格納される。この文書データベース164に格納される各文書の形式は特に限定されるものではなく、テキスト形式の文書、HTML(Hyper Text Markup Language)形式の文書、JIS形式の文書等の各種形式の文書の格納が可能である。更にこの文書データベース164には、文書を処理したユーザーのメンバー及びその処理回数が各文書に対応付けて格納されている。前記処理回数は、所定期間毎に値を0にリセットされる。文書ベクトルデータベース166には、文書データベース164に格納されている各文書に対応する文書ベクトルが格納されるようになっている。

【0012】図2は、文書ベクトルデータベース166の内容を概念的に表した説明図である。この図2に示されるように、文書ベクトルデータベース166には、上記所定期間内に処理された文書中から自動抽出されたキーワード(処理重要語(句を含む)) x 、及びこの処理重要語に対する重要度(処理重要度)が各文書の文書ベクトルの要素値 $f(x)$ として、格納されている。この文書ベクトルは各文書(A、B、C...)毎に格納され、文書データベース164に格納されている各文書と対応づけられている。

【0013】行列データベース168には、過去の所定期間に行われた文書処理の処理内容により取得される行列Ga、Gb、Gcが格納されている。GP(Group Personalize)ベクトルはこれらの行列Ga、Gb、Gcにより取得されるGP行列から取得される。図3(a)~(c)は、行列Ga、Gb、Gcの一例を示す説明図である。

【0014】行列Gaは、図3(a)に示すように、上記処理重要語を行に、同処理文書を列にとった行列であり、各要素は処理重要語の処理重要度 $f(x)$ を表している。行列Gbは、図3(b)に示すように、前記処理文書を行にとり、ユーザーのメンバー(処理者)を列にとった行列であり、各要素は、メンバーが各文書を前記

所定期間内に処理した回数となっている。この処理回数は文書データベース164から読み込まれる。行列Gcは、図3(c)に示すように、行および列がともにユーザーのメンバーそれぞれの重要度係数を示している。行列Ga及び行列Gbは所定期間ごとに書き換えられ、行列Gcは操作者からの入力により適宜書き換えられる。

【0015】記憶媒体駆動装置17は、CPU111が外部の記憶媒体からコンピュータプログラムや文書を含むデータ等を読み込むための駆動装置である。記憶媒体に記憶されているコンピュータプログラムには、本実施形態の文書処理装置により実行される各種処理のためのプログラム、および、そこで使用される辞書、データ等も含まれる。ここで、記憶媒体とは、コンピュータプログラムやデータ等が記憶される記憶媒体をいい、具体的には、フロッピーディスク、ハードディスク、磁気テープ等の磁気記憶媒体、メモリチップやICカード等の半導体記憶媒体、CD-ROMやMO、PD(相変化書換型光ディスク)等の光学的に情報が読み取られる記憶媒体、紙カードや紙テープ等の用紙(および、用紙に相当する機能を持った媒体)を用いた記憶媒体、その他各種方法でコンピュータプログラム等が記憶される記憶媒体が含まれる。本実施形態の文書処理装置において使用される記憶媒体としては、主として、CD-ROMやフロッピーディスクが使用される。記憶媒体駆動装置17は、これらの各種記憶媒体からコンピュータプログラムを読み込む他に、フロッピーディスクのような書き込み可能な記憶媒体に対してRAM113や記憶装置16に格納されているデータ等を書き込むことが可能である。

【0016】本実施形態の文書処理装置では、制御部11のCPU111が、記憶媒体駆動装置17にセットされた外部の記憶媒体からコンピュータプログラムを読み込んで、記憶装置16の各部に格納(インストール)する。そして、本実施形態による類似度算出等の各種処理を実行する場合、記憶装置16から該当プログラムをRAM113に読み込み、実行するようになっている。但し、記憶装置16からではなく、記憶媒体駆動装置17により外部の記憶媒体から直接RAM113に読み込んで実行することも可能である。また、文書処理装置によっては、本実施形態の自動要約処理プログラム等を予めROM112に記憶しておき、これをCPU111が実行するようにしてもよい。

【0017】通信制御装置18は、他のパーソナルコンピュータやワードプロセッサ等との間でテキスト形式やHTML形式等の各種形式の文書やビットマップデータ等の各種データの送受信を行うことができるようになっている。入出力I/F19は、音声や音楽等の出力を行うスピーカ等の各種機器を接続するためのインターフェースである。文字認識装置20は、用紙等に記載された文字をテキスト形式やHTML等の各種形式で認識する装置であり、イメージスキャナや文字認識プログラム等

で構成されている。

【0018】本実施形態では、キーボード12の入力操作により作成した文書(RAM113の所定格納エリアに格納)の他、外部で作成して所定の記憶媒体に格納した文書で記憶媒体駆動装置17から読み込んだ文書、予め文書データベースに格納されている文書、通信制御装置18からダウンロードした文書、及び文字認識装置20で文字認識した文書、等の各種文書を対象文書として取得する(文書取得手段)ことが可能である。

【0019】次に、上述のような構成の文書処理装置の動作であって、本発明の文書処理方法の一実施形態について図4～図9を参照して説明する。

【0020】本実施形態においては、所定期間毎に、該所定期間内に行われた文書処理の処理内容に基づいて新たな処理重要語及び処理重要度が取得され、行列データベース168内の行列Ga及び行列Gbが書き換えられる。

【0021】図4は、行列Ga、Gb書き換え処理の動作を表したフローチャートである。CPU111は、所定期間内に処理された文書(処理文書)を文書データベース164から順次取得してRAM113の所定作業領域に格納し(ステップ11)、各処理文書についての重要語(処理重要語)及びその重要度(処理重要度)を取得する(ステップ12)。

【0022】図5は、各文書についての処理重要語・処理重要度取得処理の動作を表したフローチャートである。図5に示すように、CPU111は、文書データベース164から取得した処理文書について、各処理文書毎に形態素解析を行うことで自立語を抽出する(ステップ121)と共に、名詞句、複合名詞句等を含めた候補語(句)を処理文書から抽出する(ステップ122)。次に、抽出した候補語(句)の処理文書での出現頻度、評価関数から、各候補語(句)の処理重要度 $f(x)$ を取得する(ステップ123)。ここで、評価関数としては、例えば、所定の重要語が予め指定されている場合にはその重要語に対する重み付け、単語、名詞句、複合名詞句等の候補語(句)の種類による重み付け等が使用される。

【0023】さらにCPU111は、取得した処理重要度 $f(x)$ の値をもとに候補語(句)から処理重要語a、b、c、…を取得し(ステップ124)、この処理重要語a、b、c、…及びその処理重要度 $f(a)$ 、 $f(b)$ 、 $f(c)$ …を重要語データベース165に格納する(ステップ125)。すべての処理文書について、処理重要語及びその処理重要度を取得すると、図4に示す行列Ga、Gb書き換え処理ルーチンへリターンする。

【0024】続いて、CPU111は、行列データベース168の行列Gaを、前記処理重要語a、b、c、…を行に、前記所定期間の処理文書を列に、また処理重要

度 $f(x)$ を各要素にとったものを書き換える(ステップ13)。このとき、行列 G_a の行数は、各処理文書の処理重要語の和集合の数とし、各処理文書において含まれていない処理重要語については、その処理重要度 $f(x)$ は0と定義される。

【0025】例えば図2において、処理文書Bの処理重要語は「重要、重要語、重要度、…」、処理文書Cの処理重要語は「重要、…、政治、…」であり、これらの処理重要語に対応する処理重要度は、処理文書Bについては(1, 18, 19, …)、処理文書Cについては(1, 8, …, 21, …)である。これに対して行列 G_a においては、その行は「重要、重要語、重要度、…、政治、…」とし、両文書の列における要素値はつぎの通り定義される。

処理文書Bの列=(1, 18, 19, …, 0, …)

処理文書Cの列=(18, 0, 0, …, 21, …)

【0026】また、CPU111は、文書データベース164から、各文書の処理回数を取得し(ステップ14)、行列 G_b を、所定期間内の処理文書を行に、文書データベース164から取得した処理回数を各要素としたものに書き換えて(ステップ15)、行列 G_a 、 G_b 書き換え処理を終了する。

【0027】GP行列の取得に際しては、CPU111は、前述のようにして取得され格納された行列 G_a 、 G_b 、 G_c を行列データベース168から取得し、次の式に従ってGP行列を取得する。 $GP = G_a \cdot G_b \cdot G_c$ に従って、本実施形態におけるGP行列は、文書ベクトル取得に用いられたキーワードを行に、ユーザーの各メンバーを列にとってなっており、GP行列の各要素は、メンバー毎の過去の文書処理におけるキーワードの重要度 $f(x)$ に各メンバーの重要度を加味して表した数値となっている。

【0028】続いて、本実施形態におけるGP行列の視覚化処理の動作について図6及び図7を用いて説明する。図6はGP行列の視覚化処理の動作を示すフローチャートである。GP行列が取得されると、続いてCPU111は、基準文書を取得し(ステップ21)、RAM113の対象文書格納エリア1131に格納する。基準文書は、操作者からの指示に従って、RAM113、記憶装置16の文書データベース164、記憶媒体駆動装置17、または通信制御装置18から取得する。そして、CPU111は、対象文書格納エリア1131に格納した基準文書の文書ベクトル V を求める(ステップ22)。

【0029】図7は、文書ベクトル作成処理の動作を表したフローチャートである。CPU111は、文書ベクトルデータベース166に格納されているキーワードを、基準文書から検出(ステップ221)し、基準文書での出現頻度、評価関数から、キーワードの重要度 f

(x)を得る(ステップ222)。そして、各キーワードの重要度 $f(x)$ を要素として、文書ベクトル $V=(f(a), f(b), \dots)$ を取得し(ステップ223)、RAM113の文書ベクトル格納エリア1134に格納し(ステップ224)して、図6に示すGP行列視覚化処理にリターンする。

【0030】続いて、CPU111は、文書ベクトルとGP行列との次元合わせを行う(ステップ23)。即ち、文書ベクトル V の次元数とGP行列の行数とを、基準文書のキーワードとGP行列の行があらわす処理重要語の和集合の数とし、文書ベクトル V のみに含まれるキーワードに対する行列 G_a の要素値、および、GP行列の行のみに含まれる重要語に対する文書ベクトル V の要素値は、“0”と定義する。例えば、基準文書のキーワードが「重要、重要語、重要度、…」、GP行列の行があらわす処理重要語が「重要、…、政治、…」であり、基準文書の文書ベクトル $V=(1, 18, 19, \dots)$ 、GP行列のある1列が(18, …, 21, …)である場合、次元を合わせると、基準文書の文書ベクトル $V=(1, 18, 19, \dots, 0, \dots)$ 、GP行列の1列は(18, 0, 0, …, 21, …)となる。

【0031】続いてCPU111は、次元合わせをした後のGP行列をもとにGPベクトルを取得する(ステップ24)。図8は、GP行列からGPベクトルを算出する行程を概念的に説明する説明図である。

【0032】CPU111は、まず、GP行列の各要素 g_{ij} ($i=1 \sim$ メンバー数 m 、 $j=1 \sim$ 処理重要語の和集合の数 k)の各行毎の要素の平均値を算出して列ベクトル(総GPベクトル)を得る(図8(1)→

(2))。この総GPベクトルは、各要素 g_i が処理重要語毎のユーザーグループ全体における過去の文書処理での出現頻度(但し各処理重要語の予め決められた処理重要語の重み等や、メンバーの重要度が加味されている)を反映した数値となっている。CPU111は、更に、この総GPベクトルの各要素 g_i を文書の処理回数の総数で割って、1列のGPベクトルを得る(図8(2)→(3))。この様に、総GPベクトルを文書の処理回数の総数で割るのは、行列 G_b に文書の処理回数が要素として含まれており、処理回数が増えるに従ってGPベクトルが大きくなっていくのを回避し、異なる期間の長さにおいてGPベクトルを求めても、期間の長さが影響しなくするためである。

【0033】続いて、CPU111は、そして、CPU111は、GPベクトルの各要素とこの各要素に対応する文書ベクトル V の要素とを掛け合わせて、嗜好文書ベクトル V' を得る。嗜好文書ベクトル V' は、嗜好文書ベクトルデータベース167に格納して(ステップ25)。嗜好文書ベクトル取得処理を終了する。

【0034】次に、CPU111は、文書嗜好ベクトル $V'=(f'(a), f'(b), \dots)$ の要素 f'

(a)、f' (b)、...を分野別に区分する(ステップ26)。図9は文書嗜好ベクトルV'の各要素を区分する分野の一例を示す表である。そして、分野別に要素をまとめて合計して分野別重要度F(X)を算出し(ステップ27)、分野別重要度F(X)の最も高い3分野を選択し、これらの3分野の分野別重要度F(A)、F(B)、F(C)を要素とする分野別ベクトルV'' = (F'(A)、F'(B)、F'(C))を、前記3分野をx軸、y軸、z軸とした3次元の座標上に表現して表示装置14上に表示して、GP行列の視覚化処理を終了する(ステップ28)。図10は、2つのユーザー

(A、B)それぞれの分野別ベクトルを表示装置14に表示した一例を示すものである。このように、本実施形態においては、GP行列は、分野別ベクトルV''として3次元に視覚化され表示される。この分野別ベクトル表示から、ユーザーAは、政治および環境・自然分野に嗜好が強く、ユーザーBは、ライフサイエンス分野に嗜好が強い傾向があることが一目で理解できる。

【0035】この様に、本実施形態によると、ユーザーの嗜好を表すGP行列により分野別ベクトルV''が取得され、ユーザーの嗜好の反映された分野別ベクトルV''を表示装置14に3次元表示するので、ユーザーの嗜好が目視により確認できる。

【0036】尚、本発明は、上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない限りにおいて適宜変更が可能である。例えば、上述の実施形態においては文書処理装置としてコンピュータを用いているが、コンピュータに限定されるものではなく、ワードプロセッサ等であってもよい。上述の実施形態においては、GP行列は、処理者の過去の文書処理回数(行列Ga)と各文書におけるキーワードの出現頻度(行列Gb)、および各処理者の重要度(行列Gc)とから取得されているが、処理者毎の過去の文書処理回数(行列Ga)と各文書におけるキーワードの出現頻度(行列Gb)のみにより取得してもよい。また、例えば、各文書の処理時間や、他の文書作成に引用された件数等も加味して取得してもよい。更に、GP行列を上述の実施形態と同様に行列Ga～行列Gc等の各行列から取得する場合において、行列Ga～行列Gc等の各行列の要素はそれぞれキーワードの文書中の出現頻度や、メンバーが各文書を処理した回数を反映した数値となっていればよく、直接出現頻度や処理回数そのものを表していなくてもよい。上述の実施形態においては行列Ga～Gcは所定期間毎に書き換えられているが、文書処理を行う毎に、または所定回数の文書処理を行う毎等にも書き換えてもよい。

【0037】GP行列の視覚化は、GPベクトルにより基準文書をシフトさせて取得した文書嗜好ベクトルをn次元化して表示せずに、GPベクトルを直接n次元化して表示してもよい。

【0038】また、文書嗜好ベクトルやGPベクトルの表示は、分野別ベクトルのように3次元に変換して表示しなくてもよく、例えば、図11に示すように、要素(キーワード)毎に要素値(重要度)をカラーバーで表したり、レーダーチャートにより表示する等、GPベクトルの全ての要素について表示してもよい。更に、文書嗜好ベクトルやGPベクトルを3次元に変換して表示する場合であっても、その変換手法は、上記実施形態の如く分野別に要素をまとめて合計した分野別重要度F(X)の最も高い3分野を選択した分野別ベクトルV'' = (F'(A)、F'(B)、F'(C))を表示する手法に限られるものではなく、要素を3分野に区分して分野別に要素をまとめて3次元のベクトルとする手法や、GPベクトルの要素のうちのもっとも値の高い3つを要素として3次元のベクトルとする手法等とすることもできる。文書嗜好ベクトルやGPベクトルを3次元に変換して表示する場合であっても、その表示手法は、3次元座標上にベクトルのまま表示する以外の手法でもよく、例えば、(x、y、z)軸にかえて3色(赤、

緑、青)の色を用いて各要素の値をこれらの3色の輝度に換えた色表示等で表現してもよい。上記実施形態のように3次元での文書嗜好ベクトルやGPベクトル表示する場合に、更にその軸をマウスによりポイントする等で指定すると、図12に示すように、軸が表す分野に含まれるキーワードが表示され、このキーワード中の1つをポイントすることにより操作者に選択させて当該キーワードを軸とするベクトルを表示するようにし、文書嗜好ベクトルの各要素を分野別にまとめず、各要素のうち最も値の高い3つのキーワードを軸として3次元表示してもよい。

【0039】嗜好文書ベクトルV'とともに文書ベクトルVを表示してもよい。このように嗜好文書ベクトルV'と文書ベクトルVの両方を表示することにより、ユーザーの嗜好を、文書ベクトルVと嗜好文書ベクトルV'とのなす角度として認識可能となる。一定期間毎に区切って文書嗜好ベクトルやGPベクトルを求めて、このGPベクトルの経時的変化を目視可能に表示して、ユーザーの嗜好の変化を追跡できるようにしてもよい。このように文書嗜好ベクトルやGPベクトルの経時的変化を目視可能に表示する手法としては、図13に示すように、分野別ベクトルの終点の奇跡を曲線として表示するものや、図14に示すように、カラーバグラフを重ねて表示するもの等が挙げられる。

【0040】

【発明の効果】以上説明したように、本発明によれば、ユーザーの嗜好を特徴付けるGP行列が視覚化表示されるので、ユーザーの嗜好が目視により確認できる。

【図面の簡単な説明】

【図1】本発明の文書処理装置の一実施形態であり、本発明の文書処理プログラムを記憶した記憶媒体の一実施

形態の該プログラムが読み取られたコンピュータの構成を表したブロック図である。

【図2】図1の実施形態における文書ベクトルデータベースの内容を概念的に表した説明図である。

【図3】図1の実施形態における行列Ga、Gb、Gcの一例を示す説明図である。

【図4】図1の実施形態による行列Ga、Gb書き換え処理の動作を示すフローチャートである。

【図5】図1の実施形態による処理重要語・処理重要度取得処理の動作を示すフローチャートである。

【図6】図1の実施形態によるGP行列の視覚化処理の動作を示すフローチャートである。

【図7】図1の実施形態による文書ベクトル作成処理の動作を表したフローチャートである。

【図8】図1の実施形態におけるGPベクトルのその取得手法を示す説明図である。

【図9】図1の実施形態における文書嗜好ベクトルの各要素を区分する分野の一例を示す表である。

【図10】図1の実施形態において2つのユーザーそれぞれの分野別ベクトルを表示装置に表示した一例を示すものである。

【図11】本発明の他の実施形態におけるGP行列視覚化手段のGPベクトルの表示手法の一例を示す図である。

【図12】本発明の他の実施形態におけるGP行列視覚化手段のGPベクトルの表示手法の一例を示す図である。

【図13】本発明の他の実施形態におけるGP行列視覚化手段のGPベクトルの表示手法の一例を示す図である。

【図14】本発明の他の実施形態におけるGP行列視覚化手段のGPベクトルの表示手法の一例を示す図である。

【図15】請求項1に記載した発明のクレーム対応図である。

10

【符号の説明】

11 制御部

112 ROM

113 RAM

1131 対象文書格納エリア

1132 キーワード格納エリア

1134 文書ベクトル格納エリア

1135 行列格納エリア

1136 類似度格納エリア

1138 GPベクトル格納エリア

12 キーボード

20

13 マウス

14 表示装置

15 印刷装置

16 記憶装置

161 仮名漢字変換辞書

162 プログラム格納部

164 文書データベース

165 重要語データベース

166 文書ベクトルデータベース

168 行列データベース

30

101 GP行列取得手段

102 GP行列視覚化手段

103 文書ベクトル取得手段

201 GP行列取得機能

202 GP行列視覚化機能

*

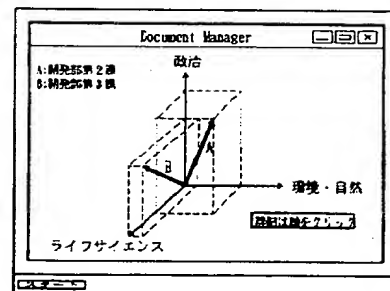
203 文書ベクトル取得機能

【図2】

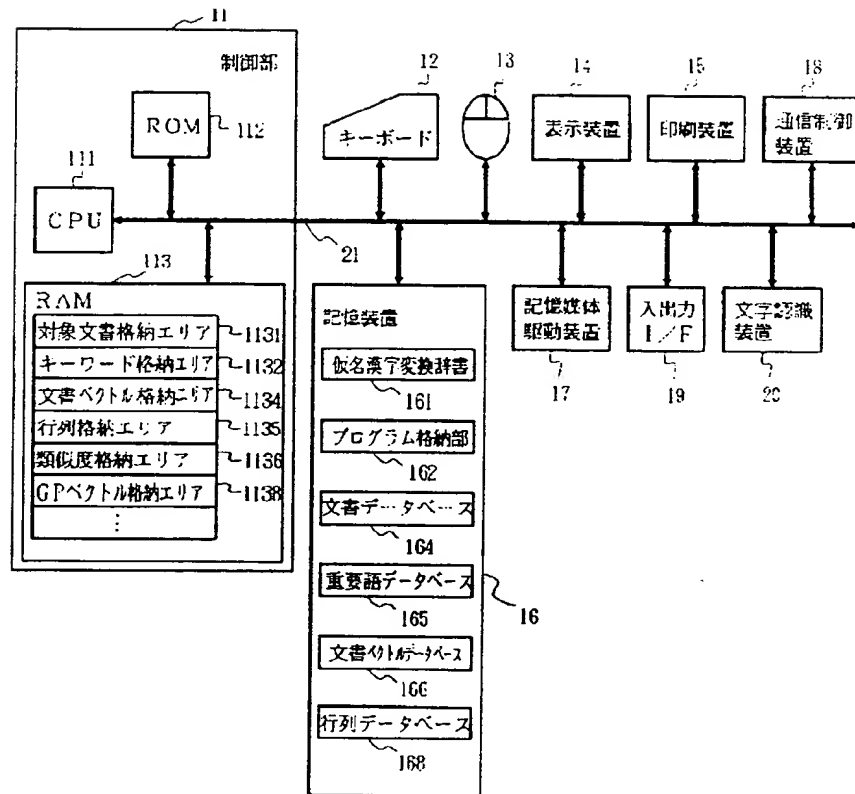
文書ベクトルデータベース

文 書	キーワードの要素値 f(x)					
	重 要	重要語	重要度	政 治	環 境・自然	ライフサイエンス
A	2	20	21	—	—	—
B	1	15	19	—	—	—
C	18	—	—	21	—	—
...						

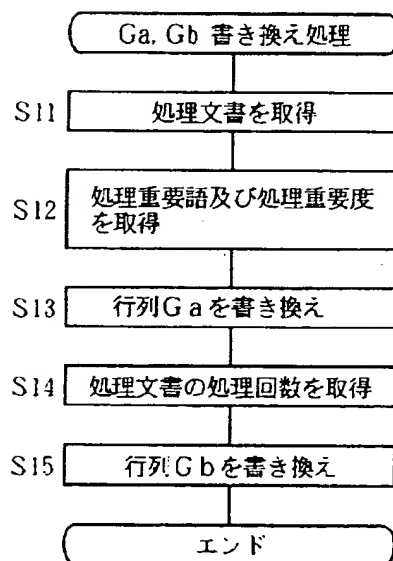
【図10】



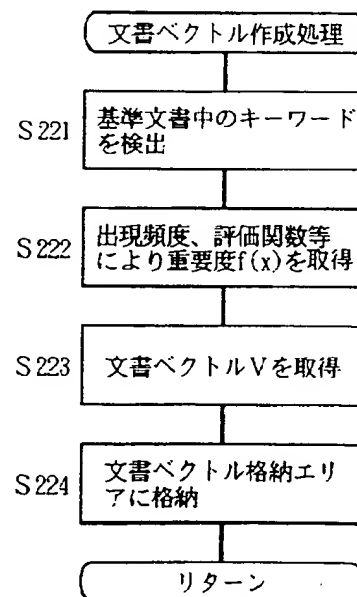
【図1】



【図4】



【図7】



【図3】

(a) 行列 G_a (キーワード、文書)

	文書A	文書B	文書C
重要語	2	1	18
重要度	20	18	0
重要度	21	19	0
政治	0	0	21

(b) 行列 G_b (文書、処理者)

	星太郎	花園英子	黒崎三四郎	鹿見五郎
文書A	1	0	1	0
文書B	1	1	2	0
文書C	1	1	1	1

(c) 行列 G_c (処理者の重要度)

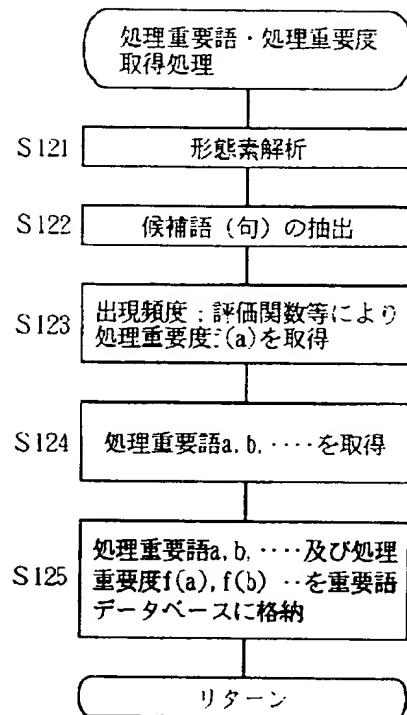
	星太郎	花園英子	黒崎三四郎	鹿見五郎
星太郎	1.5	0	0	0
花園英子	0	0.8	0	0
黒崎三四郎	0	0	1.3	0
鹿見五郎	0	0	0	1.1

【図9】

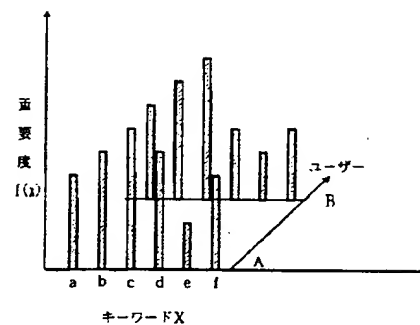
分野番号テーブル

分野	分野番号
政治 (政治・行政)	01
ライフサイエンス (医療・薬学)	02
環境・自然 (環境・気象・天文・動植物)	03
科学 (物理・化学・統計・数学)	04
工業・工学 (工業・工学・機械・電気・土木・建築)	05
文化・芸術 (映画・美術・工芸)	06
人文 (文学・言語・哲学・倫理学・心理学)	07
歴史・民俗学・考古学	08
情報・通信・コンピュータ	09
経済・経営	0a
農林水産業	0b
法律	0c
宗教	0d
スポーツ	0e
音楽	0f
料理	10
オフィス文書 (庶務・人事・経理)	11
手紙・挨拶文	12
防衛	13

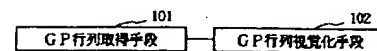
【図5】



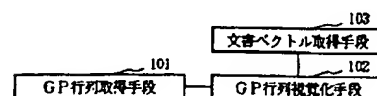
【図11】



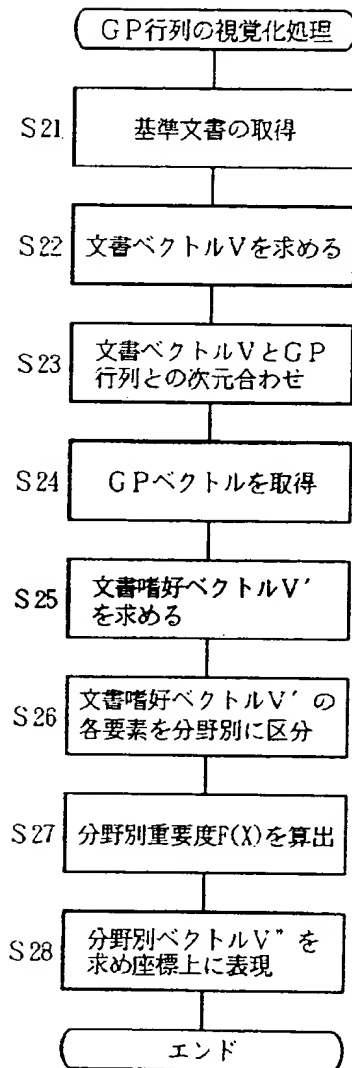
【図15】



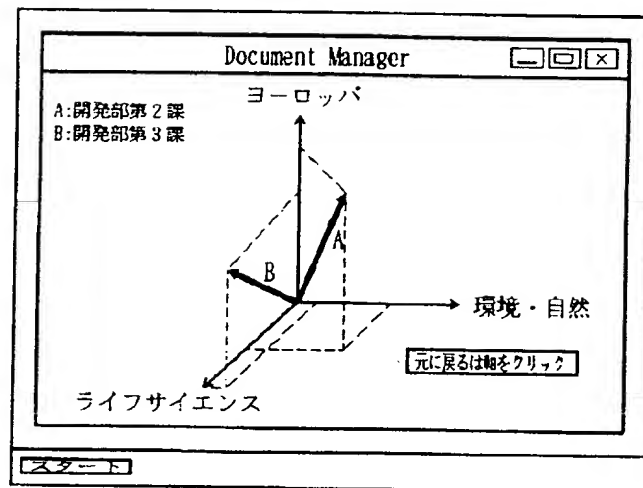
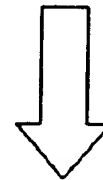
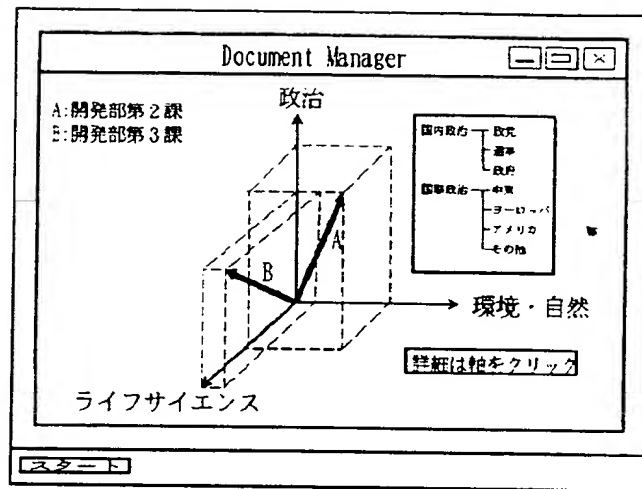
【図16】



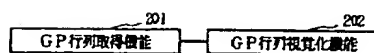
【図6】



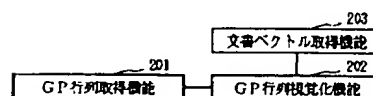
【図12】



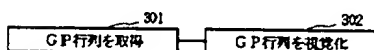
【図17】



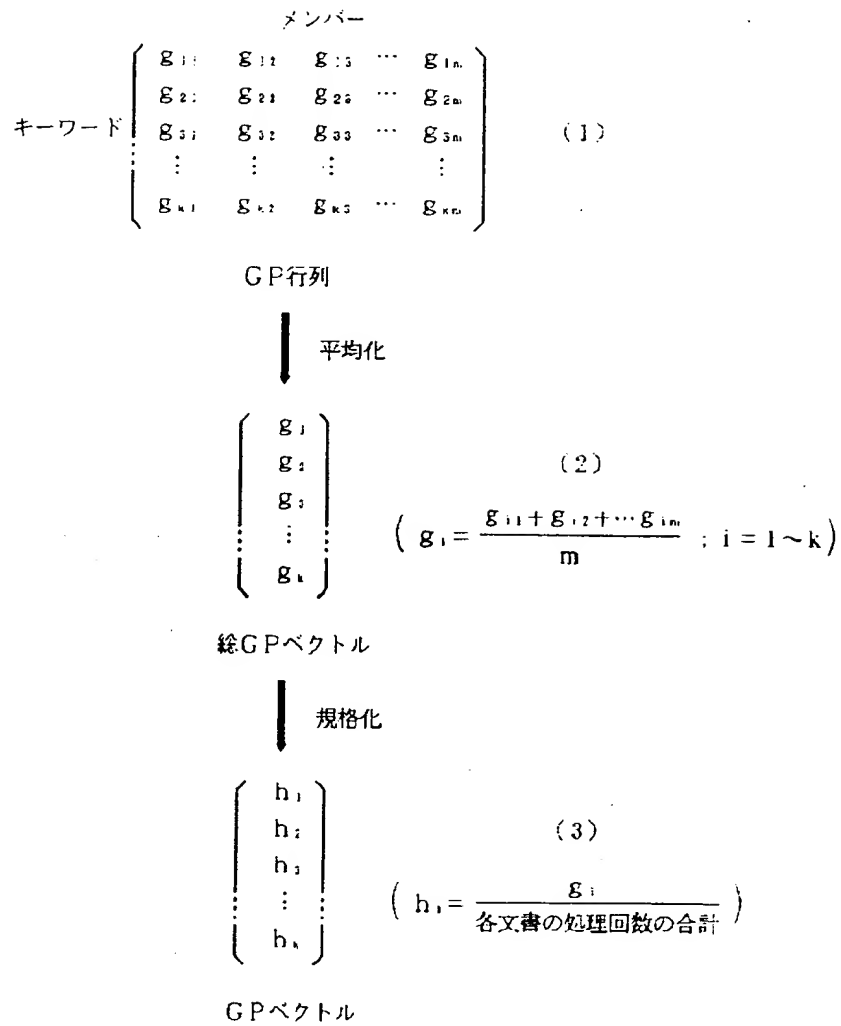
【図18】



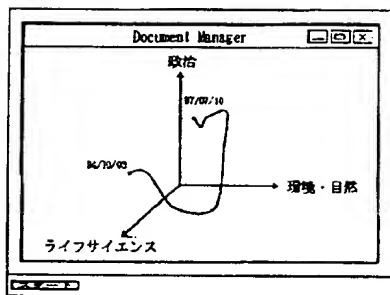
【図19】



【図8】



【図13】



【図14】

